

Leistungsparameter und Ratings in der Forschung

Hornbostel, Stefan

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Hornbostel, S. (2004). Leistungsparameter und Ratings in der Forschung. In *Evaluation – ein Bestandteil des Qualitätsmanagements an Hochschulen* (S. 173-188). Hamburg: Hochschulrektorenkonferenz (HRK). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-25347>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Leistungsparameter und Ratings in der Forschung

Dr. Stefan Hornbostel

Universität Dortmund

I. Einleitung

Während die Evaluation der Lehre inzwischen fast zur Normalität an deutschen Hochschulen geworden ist, gilt das für die Forschung nicht in gleichem Maße, obwohl die Instrumente zur Bewertung von Forschungsleistungen sehr viel früher und auch intensiver entwickelt und geprüft wurden als das Handwerkszeug der Lehrevaluation.

Die Gründe für die starke Konzentration auf die Evaluation der Lehre sind leicht nachvollziehbar: Die Lehre ist lokal gebunden, durch örtliche Praxen und Koordinationen geprägt, ohne ein eigenständiges Kommunikationssystem. Jenseits von Rahmenprüfungsordnungen blieb meist völlig im Dunklen, was sich eigentlich in den Hörsälen und Seminarräumen abspielte. Entsprechend groß war der Bedarf an Information über die Ausbildungsqualität. Für die Forschung hingegen galt schon immer, dass sie national oder international orientiert ist, institutionenübergreifend ihre Gütekriterien aus einem disziplinären Diskurs bezieht und dazu über ein eigenes Kommunikations- und Bewertungssystem verfügt, das auch eine fachliche und überfachliche Reflexion der Gütekriterien beinhaltet. Insofern war Forschung schon immer kompetitiver und transparenter als die Lehre, nicht zuletzt durch einen in den letzten Jahrzehnten erheblichen gestiegenen Anteil der drittmittelfinanzierten Forschung. Daher bestand zunächst ein eher geringes Interesse an einer vergleichenden Bewertung der Forschungsleistungen.

Das hat sich allerdings in den letzten Jahren stark verändert. Inzwischen werden die verschiedensten internationalen Forschungsrankings veröffentlicht, auf nationaler Ebene bietet das CHE ein jährliches hochschulvergleichendes Forschungsranking, die DFG legt regelmäßig ihr Förder-ranking vor und – last not least – finden auf Länderebene und auch

innerhalb der einzelnen Hochschulen aktuell sehr viele Versuche statt, Indikatorensysteme zur Bewertung von Forschungsleistungen zu etablieren – meist verbunden mit Überlegungen zu einer leistungsbezogenen Mittelverteilung. Diese Entwicklung ist in einem wettbewerblich orientierten Klima nicht überraschend, gleichwohl wirft sie nicht nur die Frage danach auf, mit welchen Indikatoren Forschungsleistungen abgebildet werden können, sondern auch danach, wie zuverlässig die verfügbaren Methoden sind.

II. Ziele und Bewertungsdimensionen

Für eine Beurteilung von Forschungsleistungen ist zunächst einmal – wie bei der Bewertung von Lehrleistungen – eine Definition von Zielen notwendig. Dazu gehören zumindest drei unterschiedliche Aspekte: Erstens die Frage danach, ob das Ausmaß der aktiven Teilnahme an der disziplinen Forschung den eigenen Zielen entspricht. Zweitens, ob die

<i>Tab. 1: Ziele Ressourcen</i>	Quantität/ Aktivität	Qualität/ Relevanz	Verwertbarkeit/ Transfer
Personal			
Finanzen			
Infrastruktur			

Qualität – oder genauer: die Resonanz – der Forschung den Erwartungen entspricht. Schließlich stellt sich, drittens, für anwendungsorientierte Forschung die Frage danach, ob die Verwertbarkeitserwartungen erfüllt werden. Diese Fragen zielen auf Effektivität (den Grad der Zielerreichung). Praktisch geht es aber nie um Effektivität um jeden Preis, vielmehr soll in der Regel der gewünschte Output mit einem möglichst geringen Ressourceneinsatz erreicht werden (Effizienz). Bereits hier sollte deutlich geworden sein, dass sich sinnvoller Weise nicht von einem Forschungsindikator sprechen lässt, sondern nur von einem Bündel unterschiedlicher Indikatoren. Allerdings betonen die verschiedenen Rankings und Ratings die Zellen der obigen Matrix auf verschiedene Weise.

III. Rating und Ranking

Die Unterschiede zwischen indikatorenbasierten Vergleichen (Ranking) und expertengestützten Bewertungen (Rating) sind inzwischen gradueller Natur. Typische Forschungsratings wie etwa das britische Research Assessment Exercise (RAE) greifen durchaus auf indikatorisierte Informationen zurück und werden von den Massenmedien in Form von Rankings (im Sinne einer Rangfolge) verbreitet. Die Gemeinsamkeiten resultieren vor allen Dingen daraus, dass eine Bewertung von Forschungsleistungen nur mit Bezug auf disziplinäre Standards möglich ist, sei es durch kompetente Experten oder durch bibliometrische Vergleichsdaten. Der Zahlenwert, den ein Indikator liefert, ist für sich genommen meist wenig aussagekräftig. Interpretierbar wird er erst durch einen Vergleich an „benchmarks“, die sich aus der Zielformulierung herleiten. Dies können vergleichbare Institutionen sein, aber auch abstrakte Größen, wie etwa die weltweit durchschnittlich erzielten Zitationswerte in einer Subdisziplin. Erst mit der Auswahl einer zielkompatiblen Referenzebene lassen sich Aussagen über die Effektivität (also den Grad der Zielerreichung) treffen. Die ursprüngliche Form des Rankings, die ausschließlich auf die Zuweisung von Rangplätzen abstellte, unabhängig davon, ob zwischen diesen Plätzen signifikante Unterschiede bestehen, ist inzwischen eher unüblich. Verbreitete Forschungsrankings wie etwa das CHE-Ranking oder das DFG-Förderranking arbeiten mit Ranggruppen und nicht mit einzelnen Rangplätzen. Davon abgesehen, greifen auch die quantifizierenden Verfahren der Forschungsevaluation immer auf die Urteile von Peers zurück. Das Grundkonzept von Wissenschaftsindikatoren basiert gerade auf der Überlegung, dass Wissenschaft nur von Wissenschaftlern bewertet werden kann. Solche Bewertungen fallen explizit als Urteile über Forschung oder Forschungsvorhaben an, implizit durch den Umgang mit den Erkenntnisansprüchen anderer.

IV. Indikatoren

Urteile über die Güte, Nutzbarkeit oder Relevanz von Forschung entstehen also in der Kommunikation über Forschung. Die Urteile, die in diesem sozialen Raum gefällt werden, sind nicht mit einer methodologischen Würdigung der Forschungsbefunde zu verwechseln. Die Güte von Forschung besteht in einem methodologischem Sinne nur bedingt aus Ak-

zeptanz und Nutzung. Forschungsindikatoren sind daher Konstruktionen, die bestenfalls das wiedergeben, was Wissenschaftler für gut, wichtig oder nützlich halten.

1. Peer-Urteile

Zu den routinemäßig anfallenden Peer-Urteilen gehören die Begutachtung von Drittmittelanträgen, die Manuskriptbeurteilung bei den Zeitschriften, Jurorenurteile bei Preisverleihungen und bedingt auch Reputationsurteile, die in Befragungen erhoben werden. All diese Urteile stellen ausschnittshafte Bewertungen der Forschungsaktivität dar, die sich mehr oder weniger gut für die Beschreibung der Forschungsperformanz ganzer Institutionen eignen. Die Nutzung von Angaben über die Höhe der eingeworbenen Drittmittel als Indikator für Forschungsqualität setzt voraus, dass Drittmittelforschung in der untersuchten Disziplin üblich ist. Wie die folgende Grafik zeigt, ist das keineswegs in allen Disziplinen gegeben.

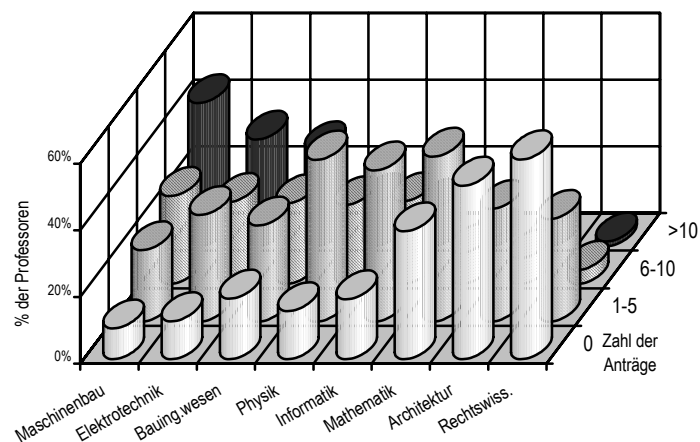


Abb.1: Drittmittelanträge von Universitätsprofessoren in den letzten 3 Jahren (Prozent der befragten Professoren)

Quelle: Hornbostel, CHE Studienführer 1999 und 2000; Erhebungszeitraum: Maschinenbau, Elektrotechnik, Architektur 1999, alle anderen 1998

Juristen werben kaum Drittmittel ein, während bei den Maschinenbauern die Mehrzahl der Professoren jährlich mehrere Anträge stellt. Im Falle der Juristen würde eine Nutzung des Drittmittelindikators äußerst selektiv über einen bestimmten Typ von Forschung berichtet werden. Im Fall der Maschinenbauer stellt sich die Frage, ob auch all jene Mittel, die im Rahmen von Entwicklungs- und Anwendungsaufträgen akquiriert wurden, als Indikatoren für Forschungsaktivität tauglich sind. Denn Drittmittelerwerbungen indizieren Forschungsqualität nur in so weit, als die Annonce eines Forschungsvorhabens die kritische Begutachtung durch Fachwissenschaftler übersteht.

Wissenschaftspreise gehören sicherlich zu den expliziten Urteilen über die Forschungsqualität. Für die Indikatorenkonstruktion werden sie jedoch nur selten genutzt, da erstens der größte Teil der Wissenschaftspreise hochspezialisiert ist und zweitens der Zwang zur Auswahl unter vielen ähnlich qualifizierten Arbeiten eine Graduierung, wie sie für Indikatoren notwendig ist, verhindert.

Umgekehrt finden sich Reputationserhebungen – nicht nur in deutschen Forschungsrankings – relativ häufig wieder, was unter anderem mit der leichten Erhebbarkeit derartiger Daten zu tun hat. Meist entsteht dabei ein weitaus stärkeres Gefälle zwischen den sehr renommierten Institutionen und den weniger bekannten, als auf Leistungsindikatoren sichtbar wird. Methodisch ist dies auf sog. „Star-“ und „Hello-Effekte“ zurückzuführen. Problematischer aber ist, dass Reputation nur sehr langsam erworben wird und ebenso langsam verloren geht. Newcomer haben es daher sehr schwer, eine angemessene Aufmerksamkeit zu erreichen. Unterhalb der besonders bekannten Institutionen führen Reputationserhebungen meist zu wenig validen Positionierungen. Der Wert von Reputationserhebungen liegt eher darin, dass sie eine Art Realität *sui generis* beschreiben, die sich von Leistungsbezügen durchaus entkoppeln kann, aber unter Marketinggesichtspunkten dennoch Bedeutung hat. Auch die Analyse von Stipendiatenzahlen, die sich für einen Aufenthalt an einer deutschen Forschungseinrichtung bewerben, folgt der Logik von Reputationserhebungen.

Die Manuskriptbeurteilung wird bei der Konstruktion von Forschungsindikatoren dazu genutzt, um Zeitschriften als „reviewed“ zu klassifizieren. Aus den Bibliographien der Wissenschaftler werden dann nur noch jene Zeitschriftenartikel berücksichtigt, die in einem Journal mit Reviewverfahren erschienen sind. Allerdings existieren keine Standards dafür, wann ein Reviewverfahren eine hinreichende Dignität aufweist. Bei weitem nicht alle wissenschaftlichen Zeitschriften benutzen ein „double-blind“ Verfahren, um eine möglichst verzerrungsfreie Beurteilung zu garantieren. Bis vor kurzem haben darüber hinaus auch angesehene Zeitschriften allein durch die Herausgeber über Annahme oder Ablehnung von Manuskripten entschieden. Gelegentlich wird auch versucht die Güte des Reviewverfahrens durch Befragungen unter den Fachwissenschaftlern zu ermitteln. Die Schwachstelle einer solchen Klassifikation ist vor allen Dingen, dass sie der Qualität des einzelnen Beitrages durch die globale Einstufung in der Regel nicht gerecht wird. Charakterisierungen von Forschungsbeiträgen durch die Zeitschrift, in der sie erschienen sind, stellen daher eher ein Hilfsmittel für den Fall dar, dass elaboriertere Verfahren nicht anwendbar sind.

2. Verhalten der Scientific Community

Zu solchen ausgefeilteren Verfahren gehört die ganze Gruppe der bibliometrischen Techniken. Bei diesen Verfahren stehen nicht besonders erfahrene Peers im Vordergrund, sondern die Verhaltensweisen der jeweiligen Scientific Community, und zwar so, wie sie sich in der Forschungskommunikation abbilden. Die Grundannahme besteht darin, dass wichtige Forschungsergebnisse eine höhere Resonanz in der Community erzeugen als unwichtige. Ein solches Abbild der Kommunikation über Forschungsbefunde ist aber nicht einfach vorhanden, sondern muss verfahrensmäßig konstruiert werden. Das heißt zum einen, dass ein Korpus forschungsrelevanter Literatur definiert werden muss, zum anderen, dass die Bezugnahmen der Wissenschaftler auf einander rekonstruiert werden müssen. Auf die Details solcher Analysen kann hier nicht eingegangen werden. Erwähnt sei aber zumindest die disziplinär sehr unterschiedliche Eignung derartiger bibliometrischer Verfahren. Für naturwissenschaftliche Disziplinen, in denen die Forschungskommunikation nicht nur stark internationalisiert ist, im Wesentlichen durch Artikel in wissenschaftlichen

Zeitschriften getragen wird und hinsichtlich der Zeitschriften starke Reputationshierarchien aufweist, sind die Voraussetzungen für bibliometrische Analysen vergleichsweise gut. Mit dem Science Citation Index (SCI) existiert für diese Disziplinen zudem eine Literaturlatenbank, für die weltweit Zeitschriften nach Relevanzgesichtspunkten (Zitationshäufigkeit) ausgewertet werden. Für geistes- und sozialwissenschaftliche Disziplinen gilt hingegen fast das Gegenteil. Neben Zeitschriftenartikeln wird die Forschungskommunikation dort wesentlich über Monographien und Sammelbände getragen, sie sind eher national als international ausgerichtet und eine klare Hierarchisierung der Publikationsmedien fehlt meist. Tabelle 2 zeigt exemplarisch (hier Soziologie) die von deutschen Professoren bevorzugten Publikationstypen.

Tab.2: Publikationen 1998-2000 von Soziologieprofessoren an Fachbereichen mit grundständigem Lehrangebot	
	in Prozent
Zeitschriftenaufsatz	30,3
Sammelband	58,9
Monographie	5,1
Graue Literatur	5,7
Quelle: Solis/Foris, IZ-Bonn, eigene Berechnungen	

Dass Sammelbände und Monographien bei einer seriösen Analyse nicht vernachlässigt werden dürfen, zeigt sich, wenn man danach fragt, was in den führenden soziologischen Fachzeitschriften zitiert wird. Wie in oben stehender Tabelle zu erkennen ist, bezieht sich mehr als die Hälfte aller Quellenangaben auf eben jene Sammelbände und Monographien. Für die Ermittlung derartiger Daten eignet sich der SCI nicht, man ist vielmehr auf nationale Fachdatenbanken angewiesen.

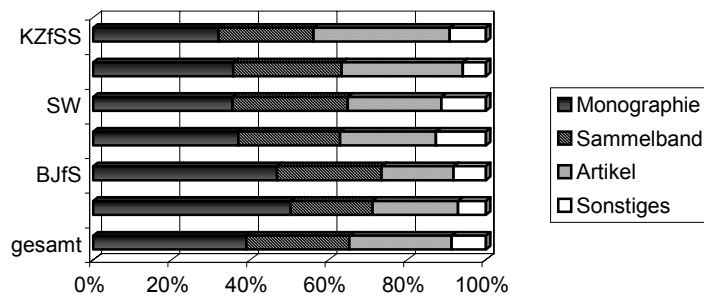


Abb. 2: Zitierte Literatur aus sechs deutschen soziologischen Fachzeitschriften des Jahrganges 2002 nach Art der zitierten Quelle

Solche Fachdatenbanken enthalten allerdings auch eine Fülle nicht einschlägiger Literatur (z.B. Graue Literatur). Das Fehlen jeglicher Relevanzmarkierungen führt dazu, dass derartige Datenbanken nur nach aufwändigen Bereinigungen nutzbar sind. Auch dann fehlt einer solchen Datenbasis allerdings der internationale Bezug. Das führt schnell zu einem Dilemma, wenn eine Disziplin sowohl nationale wie internationale Forschungslinien enthält: Für die deutsche VWL ermittelte z.B. das CHE in den Jahren 1998 bis 2000 insgesamt 4149 Publikationen von Professoren der VWL. Legt man die Kriterien eines Rankings für die europäischen Wirtschaftswissenschaften zu Grunde (Kalaitzidakis 1999), das lediglich 10 europäische Spitzenjournals (mit einer Qualitätsgewichtung) berücksichtigt, dann bleiben für die gesamte deutsche Volkswirtschaft gerade einmal 32 Publikationen in diesem Zeitraum übrig, also weniger als ein Prozent der deutschen Publikationen). Es ist kaum verwunderlich, dass beide Datenbasen zu unterschiedlichen Rankingergebnissen führen.

Während in diesem Fall unterschiedliche Datenbasen und die Verwendung unterschiedlicher Qualitätsgewichtungen zu gravierenden Differenzen führen, ergeben sich bei eher homogenen Ausgangsbedingungen weit weniger dramatische Veränderungen der Rankingposition. Abbildung 3 zeigt am Beispiel der Psychologie (Basis: Zeitschriftenartikel im SCI), wie sich die Position der Fachbereiche verändert, wenn an Stelle der

einfachen Publikationszählung eine Gewichtung der Publikationen mit der Zahl der Co-Autoren und der Anzahl der erhaltenen Zitate vorgenommen wird. Während die Gewichtung auf der individuellen Ebene zu erheblichen Veränderungen führt, verändert sich die Position der Fachbereiche vergleichsweise moderat. Der Grund liegt darin, und dass sich die Fachbereiche hinsichtlich der Komposition von viel und wenig zitierten Autoren und Artikeln recht ähnlich sind.

**Psychologie Institute nach gew. und ungew.
Publikationen (2000-2002) der Professoren
(Absolutangaben)**

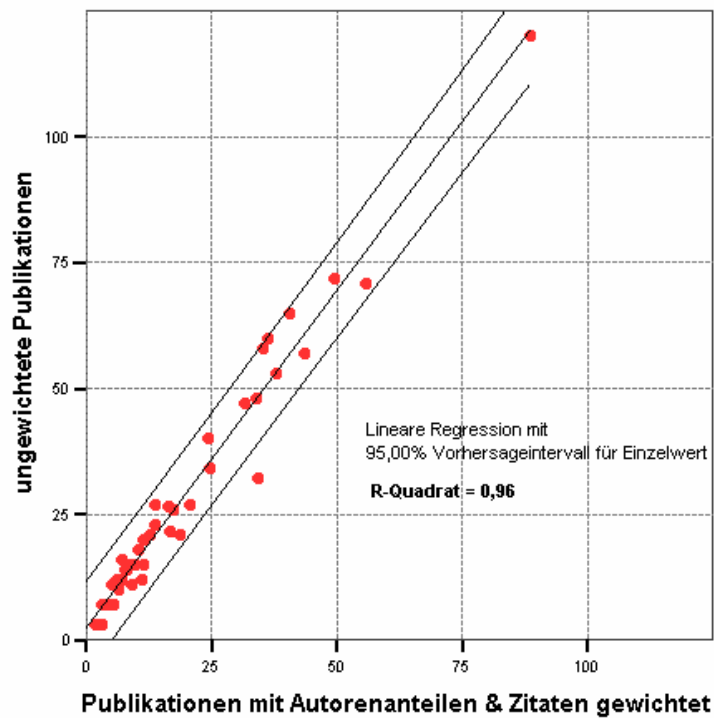


Abb. 3

Auch dieser Befund gilt allerdings nicht für alle Disziplinen gleichermaßen. Immer dann, wenn sich subdisziplinäre Forschungsfelder mit eignen Publikations- und Zitationsmustern herausgebildet haben, können bibliometrische Angaben schnell zu Missverständnissen führen. Abbildung 4 zeigt einen solchen Fall in der Physik. Bei dem „Ausreißer-Fachbereich“ mit der höchsten Anzahl von Zitaten je Artikel handelt es sich um einen kleinen Fachbereich, der sich auf ein Spezialgebiet der Physik konzentriert hat (Hochenergiephysik). Die hohe Anzahl von Zitaten ist den Gepflogenheiten dieses Spezialgebietes geschuldet und nicht einer besonders guten Performanz des Fachbereiches. Methodisch versucht man derartige Effekte durch eine Normalisierung an den Durchschnittswerten des jeweiligen Spezialgebietes (relativer Zitationserfolg) zu neutralisieren.

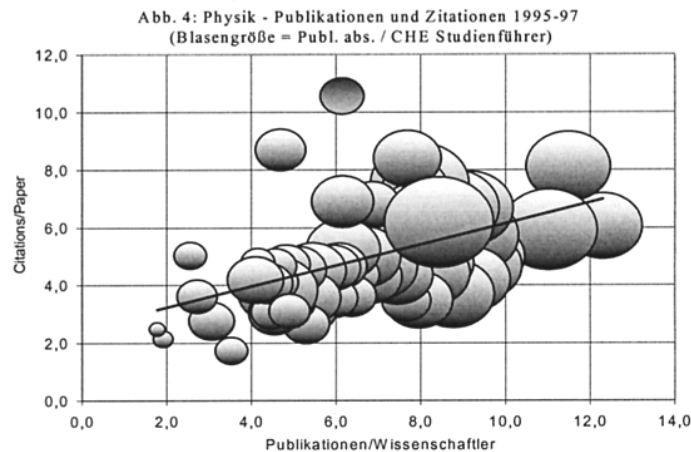


Abb. 4

Die knappe Skizze zeigt, dass bibliometrische Techniken zwar sehr mächtige Verfahren sind, die neben Performanzangaben (aktive Teilhabe am Forschungsgeschehen, Erfolg als Resonanz) auch strukturelle Informationen über Personen-, Institutionen- und Themennetzwerke liefern können, aber es handelt sich nicht um eine Routineberichterstattung, sondern um sorgfältig an die disziplinären Bedingungen anzupassende Werkzeuge.

Die Quantifizierung von Leistungen und Erfolgen durch Indikatoren erzeugt regelmäßig die Illusion besonderer Exaktheit der Beurteilung. Die korrekte Messung eines Leistungsparameters ist aber nur ein Teilaspekt bei der Verwendung von Indikatoren. Während Experten üblicherweise über den Zweck und die Kontextbedingungen einer evaluativen Stellungnahme sehr genau informiert sind und dieses Wissen in ihrem Urteil berücksichtigen, sind Indikatoren kontextfrei. Genauer gesagt: Die Kontextualisierung von Indikatoren ist Teil des Verwendungsprozesses und muss beim Einsatz von Indikatoren berücksichtigt bzw. hergestellt werden.

3. Reaktionen wissenschaftsexterner Auditorien

Im Falle anwendungsorientierter Disziplinen helfen bibliometrische Indikatoren nur bedingt weiter, denn solche Disziplinen adressieren neben einem akademischen Auditorium auch eine wissenschaftsexterne Öffentlichkeit. Indikatoren für Akzeptanz und Transfer in vor allen Dingen ökonomischen Anwendungsfeldern sind beispielsweise Industriekooperationen oder Patentanmeldungen. Zwar hat sich die Auswertbarkeit von Patentdaten seit Wegfall des Professorenprivilegs stark verbessert, auch hier gilt aber, dass die Anmeldung von Patenten nicht ohne weiteres als Indikator für Innovation verstanden werden kann. Ob Patente angemeldet werden oder nicht hängt jenseits des Innovationsgehaltes von sehr vielen Faktoren ab. Wichtiger noch: die Anmeldung von Patenten ist keineswegs gleichzusetzen mit einer tatsächlichen ökonomischen Verwertung.

4. Wissenschaftlicher Nachwuchs

In den meisten Rankings findet man unter dem Titel „Forschung“ auch Angaben zu Promotionen, Habilitationen oder Graduiertenkollegs. Genau genommen handelt es sich dabei nicht um Forschungsindikatoren, sondern um Kennzahlen, die die Ausbildung wissenschaftlichen Nachwuchses beschreiben. In Disziplinen, in denen Promotionen im Kontext von Forschungsprojekten angefertigt werden, korreliert die Anzahl der Promotionen erwartungsgemäß mit anderen Forschungsindikatoren. In Fachgebieten, in denen der Zusammenhang zwischen Projektforschung und Promotion weniger eng ist, taugen Promotionszahlen als For-

schungsindikator jedoch nicht. Die folgende Tabelle zeigt für die Soziologie, dass zwischen der Anzahl der Promotionen und den eingeworbenen Drittmitteln bzw. den Publikationen keine Zusammenhänge bestehen. Auch die Reputation der Fachbereiche (hier als Empfehlung einer Universität durch die Professoren gemessen) korreliert zwar moderat, aber immerhin signifikant mit den Drittmittelinwerbungen und den Publikationen, nicht jedoch mit der Anzahl der Promotionen.

Tab. 3: Soziologie Fachbereiche 1999-2001			
Person's Corr.			
	Drittmittel	Promotionen	Publikationen
Professorentipp (Reputation)	0,320* N=40	0,102 N=38	0,357* N=44
Drittmittel		-0,181 N=36	0,147 N=39
Promotionen			-0,146 N=35
* Korrelation ist auf 0,05 Niveau signifikant			
Quelle: CHE Forschungsranking 2002			

V. Untersuchungseinheiten

Forschung ist in ein doppeltes Bezugssystem eingebunden: zum einen gehören die Forscher organisatorischen Einheiten wie Hochschulen, außeruniversitären Einrichtungen, Instituten oder Fachbereichen (gelegentlich auch in Form von Doppelmitgliedschaften) an, zum anderen sind sie Mitglieder einer davon unabhängigen disziplinären Kommunikationsgemeinschaft. Insbesondere wenn es um Effizienzberechnungen geht, kann diese doppelte Zugehörigkeit schnell zu Verwirrungen führen. Erfolgt die fachliche Zuordnung von Personen aufgrund ihrer Organisationszugehörigkeit, dann wird z.B. der Soziologe an einer juristischen Fakultät als Jurist gezählt, der Jurist an einer medizinischen Fakultät hingegen als Mediziner. Folgt man den Fachklassifikationen des SCI, dann wird nicht die Person, sondern ein einzelner Artikel fachlich zugeordnet (d.h. eine Person wird möglicherweise mehreren Fachgebieten zugeordnet), folgt man den institutionellen Adressen im SCI, dann wird eine Person derjenigen Institution zugerechnet, deren Adresse der Autor

als Korrespondenzadresse in seinem Artikel angegeben hat („work done at“), amtliche Statistik oder DFG Fachausschüsse erzeugen wiederum andere fachliche Aggregate. Und natürlich können auch kleinere Einheiten wie Forschungsgruppen, Teilnehmer an bestimmten Programmen etc. Grundlage für eine Zuordnung sein. Selbst wenn eine fachliche Zuordnung festgelegt ist, bleibt offen, welche Personen in die Analyse einbezogen werden. Das CHE analysiert alle Professoren, andere Studien schließen den – schwer abgrenzbaren – akademischen Mittelbau ein. Ganz andere Wege geht das britische RAE. Dort werden nur Personen berücksichtigt, die als forschungsaktiv für das Evaluationsverfahren gemeldet werden – eine Differenzierung zwischen Hochschulforschern und -lehrern, die in Deutschland so kaum vorstellbar ist. Keine dieser Entscheidungen ist richtig oder falsch, sie sind nur bei gegebener Zielsetzung mehr oder weniger angemessen. In jedem Fall aber können sehr unterschiedliche Befunde entstehen, je nach Zuschnitt des Personenaggregats, das der Analyse zugrunde liegt.

VI. Wie viel Übereinstimmung muss sein?

Die skizzierten Indikatoren messen unterschiedliche Facetten des Forschungsgeschehens und daher sollte – angesichts der Komplexität des Gegenstandsbereiches – ein mehrdimensionales Indikatorensystem auch jedem Ranking zugrunde liegen. Allerdings sind ähnlich wie bei sozialen oder ökonomischen Indikatorensystemen keine völligen Übereinstimmungen zwischen den verschiedenen Indikatoren zu erwarten. Grundsätzlich lässt sich aber eine hohe Übereinstimmung verschiedener Indikatoren als Hinweis auf eine befriedigende Validität und Reliabilität der Messungen deuten. Gleichwohl kann das Auseinanderfallen verschiedener Indikatoren auch nachvollziehbare Gründe in der jeweiligen Disziplin haben. Je unterschiedlicher die Forschungslinien einer Disziplin, je heterogener die epistemischen Bezüge, je variantenreicher die empirische Basis ist und je ausgeprägter institutionelle Profile sind, desto geringer fällt vermutlich die Übereinstimmung der verschiedenen Indikatoren aus. Tabelle 4 zeigt für die Erziehungswissenschaft (für die diese heterogenen innerdisziplinären Bedingungen zutreffen) die nur sehr Korrelationen zwischen Drittmiteleinwerbungen und Publikationsaktivitäten (vgl. S. 186, Abb. 5).

Tab. 4: Korrelationen zwischen Publikationen und Drittmiteleinwerbungen Erziehungswissenschaftliche Fachbereiche zwischen 1997 und 2002						
	Gewichtete Publikationen je Professor		Professoren am Fachbereich		Gewichtete Publikationen (absolut)	
	97- 99	00- 02	97- 99	00- 02	97- 99	00- 02
Drittmittel (insg.) je Professor	.115	.204				
Drittmittel (insg.)			.353 *	.463 **	.368 *	.565 **
DFG Mittel					.522 **	.414 **
N (Fachbereiche)	34	43	34	43	34	43
Anm: Nur Fachbereiche mit gemeldeten Drittmiteleinkünften und mehr als drei Professoren am Fachbereich. Nur Publikationen der Professoren. Für den Zeitraum 2000-2002 sind Publikationen der Jahre 1999 miterfasst. Drittmittelangaben beziehen sich auf verausgabte Mittel in den Jahren 1997-1999 resp. 2000-2002. Quelle: CHE-Studienführer 2001, 2004						

Abb. 5

Die Tabelle zeigt allerdings auch, dass die intensive innerfachliche Diskussion anhand von Forschungsrankings Wirkungen in der Disziplin entfaltet, denn in der zeitlichen Entwicklung zeigen die steigenden Korrelationskoeffizienten in Richtung höherer Übereinstimmung (es bleibt aller-

dings abzuwarten, ob hier tatsächlich substanzielle Veränderungen abgebildet werden oder die Effekte einer verbesserten Datenbasis geschuldet sind).

VII. Resümee

Forschungsevaluation entwickelt sich zum wesentlichen Bestandteil eines profilorientierten Qualitätsmanagements in der Hochschule. Die Instrumentarien für eine solche Evaluation sind zwar deutlich besser entwickelt als im Bereich der Lehrevaluation, aber dennoch weit entfernt von einem routinemäßig einsetzbaren Verfahren. Das hat einerseits damit zu tun, dass gerade komplexe Forschungsindikatoren einer kompetenten Interpretation bedürfen, andererseits damit, dass Forschungsevaluation immer auf disziplinäre Vergleichsdaten angewiesen ist. Verfahren des „informed Peer Review“ verbinden derartige indikatorisierte Berichterstattungen mit fachlicher Interpretationskompetenz.

Im Rahmen einer institutionellen Handlungsstrategie sind mehrdimensionale, auf die institutionellen Ziele und die disziplinären Bedingungen zugeschnittene Berichtssysteme sicherlich fruchtbarer, als extrem simplifizierte Kennzahlensysteme, die zwar leicht zu administrieren, aber in ihren Wirkungen nur schwer kalkulierbar sind.

In jedem Fall aber werden Hochschulen ebenso wie Fachgesellschaften in Zukunft wichtige Parameter der Forschungsaktivität, über die auch öffentlich berichtet wird, in systematischer Weise verfolgen müssen.